

PCR/US91/11016
08/324001

> 0 (<
01 10 IntelliGenetics
> 0 (<

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file jan11a.res made by on Wed 11 Jan 95 12:32:38-PST.

Query sequence being compared: CL16 (1-21)
Number of sequences searched: 302507
Number of scores above cutoff: 4620

Results of the initial comparison of CL16 (1-21) with:
Data bank : EMBL-NEW 10, all entries
Data bank : GenBank 85, all entries
Data bank : GenBank-NEW 10, all entries
Data bank : HIV-NA 7, all entries
Data bank : Issued_NA , all entries
Data bank : N-GeneSeq 16.3, all entries
Data bank : UEMBL 40_85, all entries
Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		
Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	6	7	2.92

Times:	CPU	Total Elapsed
	00:13:01.04	00:13:36.00

Number of residues: 276734581
Number of sequences searched: 302507
Number of scores above cutoff: 4620

Cut-off raised to 4.
Cut-off raised to 5.
Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.
Cut-off raised to 10.

Cut-off raised to 11.
 Cut-off raised to 12.
 Cut-off raised to 13.
 Cut-off raised to 14.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 3 standard deviations above mean ****						
1. XELRGASBC	x.borealis somatic 5s rrna ge	375	17	20	3.77	0
2. XELRGASBA	X.borealis somatic 5S rRNA ge	858	17	20	3.77	0
3. XBRNA2	Xenopus borealis gene for 5S	858	17	20	3.77	0
4. RATMAP1A	Rat MAP-1 gene encoding major	1402	17	17	3.77	0
5. RATTKG1	Rat T-kininogen (T-KG) gene,	1903	17	17	3.77	0
6. RNAMDX23	R.norvegicus S-adenosylmethio	2021	17	17	3.77	0
7. ZDHRGP	Z.diploperennis gene for hydr	4478	17	17	3.77	0
8. MMT1CPS	Mouse Tla region Tlc pseudoge	8147	17	17	3.77	0
9. RATSADMEDC	Rat AdoMetDC gene, complete C	17167	17	17	3.77	0
10. CELF28F5	Caenorhabditis elegans cosmid	32903	17	17	3.77	0
11. CELF28F5	Caenorhabditis elegans cosmid	32903	17	17	3.77	0
12. CHNTXX	Tobacco chloroplast genome DN	155844	17	17	3.77	0
13. N60861	Fragment of plasmid PXC204 en	146	16	16	3.43	0
14. T24747	EST322 Homo sapiens cDNA clon	186	16	16	3.43	0
15. HS7476	EST322 Homo sapiens cDNA clon	186	16	16	3.43	0
16. NVIRGAA	Newt (Notophthalmus viridesce	235	16	18	3.43	0
17. NV5SRRN	Notophthalmus viridescens 5S	235	16	18	3.43	0
18. N60862	Fragment of plasmid PXC204 en	288	16	16	3.43	0
19. XELCRLB	Xenopus laevis caerulein prec	301	16	16	3.43	0
20. PABKTANT	BK virus 5' end of early regi	332	16	16	3.43	0
21. XLCAER1	Xenopus laevis mRNA fragment	370	16	16	3.43	0
22. XELCRLA	Xenopus laevis caerulein prec	370	16	16	3.43	0
23. T08475	EST06366 Homo sapiens cDNA cl	383	16	16	3.43	0
24. XELCRLG35	Xenopus laevis caerulein type	391	16	16	3.43	0
25. XELCRLI	X.laevis caerulein mRNA, clon	395	16	16	3.43	0
26. N60858	Sequence of plasmid PXC102 en	397	16	16	3.43	0
27. PVBRESWW	Human papovavirus BK (strain	426	16	16	3.43	0
28. HUMUT2361	Human STS UT2361.	446	16	16	3.43	0
29. BRRRPL37A	Brassica rapa ribosomal prote	446	16	16	3.43	0
30. N50145	Sequence of enhancer DNA segm	451	16	16	3.43	0

Query sequence being compared: CL16 (1-21)
 Number of sequences optimized: 4620

Results of the optimized comparison of CL16 (1-21) with:
 Data bank : EMBL-NEW 10, all entries
 Data bank : GenBank 85, all entries
 Data bank : GenBank-NEW 10, all entries
 Data bank : HIV-NA 7, all entries
 Data bank : Issued_NA , all entries

Data bank : N-GeneSeq 16.3, all entries
 Data bank : UEMBL 40_85, all entries
 Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		

Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	14	15	0.80

Times:	CPU	Total Elapsed
	00:01:06.99	00:01:38.00

Number of residues:	24455601
Number of sequences optimized:	4620

The scores below are sorted by optimized score.
 Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 7 standard deviations above mean ****						
1. XELRGASBC	x.borealis somatic 5s rrna ge	375	17	20	7.51	0
2. XELRGASBA	X.borealis somatic 5S rRNA ge	858	17	20	7.51	0
3. XBRNA2	Xenopus borealis gene for 5S	858	17	20	7.51	0
**** 6 standard deviations above mean ****						
4. XELRGASL	X.laevis somatic 5S rRNA gene	888	16	19	6.26	0
**** 5 standard deviations above mean ****						
5. NV5SRRN	Notophthalmus viridescens 5S	235	16	18	5.01	0
6. NVIRGAA	Newt (Notophthalmus viridesc	235	16	18	5.01	0
7. XELRGAQB	x.borealis oocyte 5s dna.	761	15	18	5.01	0
8. XBRNA1	Xenopus borealis genes (three	761	15	18	5.01	0
**** 3 standard deviations above mean ****						
9. MMT1CPS	Mouse T1a region T1c pseudoge	8147	17	17	3.76	0
10. ZDHRGP	Z.diploperennis gene for hydr	4478	17	17	3.76	0
11. RATTKG1	Rat T-kininogen (T-KG) gene,	1903	17	17	3.76	0
12. CELF28F5	Caenorhabditis elegans cosmid	32903	17	17	3.76	0
13. RNAMDX23	R.norvegicus S-adenosylmethio	2021	17	17	3.76	0

14.	CELF28F5	Caenorhabditis elegans cosmid	32903	17	17	3.76	0
15.	RATMAP1A	Rat MAP-1 gene encoding major	1402	17	17	3.76	0
16.	CHNTXX	Tobacco chloroplast genome DN	155844	17	17	3.76	0
17.	RATSADMEDC	Rat AdoMetDC gene, complete C	17167	17	17	3.76	0
18.	ONHGHCOHO	Oncorhynchus kisutch (coho sa	1201	15	17	3.76	0
19.	STREIPEPA	Streptococcus salivarius phos	2259	15	17	3.76	0
20.	XLXK70A	Xenopus laevis XK70A gene for	6266	15	17	3.76	0
21.	CEF54C8	Caenorhabditis elegans cosmid	23000	15	17	3.76	0
22.	CEF54C8	Caenorhabditis elegans cosmid	23000	15	17	3.76	0
23.	MIOACYTB	O.aries mitochondrion cytb ge	1140	13	17	3.76	0
24.	SV4EV211	SV40 variant genome ev-2114,	100	14	17	3.76	0
25.	DRO7DC14Z	Drosophila melanogaster (subc	3625	13	17	3.76	0
26.	T10577	hbc220 Homo sapiens cDNA clon	560	13	17	3.76	0
27.	NEUFRG	Neurospora crassa mRNA sequen	4631	13	17	3.76	0
28.	HSA26A071	H. sapiens partial cDNA seque	347	13	17	3.76	0
29.	HSA39H101	H. sapiens partial cDNA seque	345	12	17	3.76	0
30.	PCT-US93-04648-1	Sequence 15, Application	10596	14	17	3.76	0

1. CL16 (1-21)

XELRGASBC x.borealis somatic 5s rrna gene, clone pxbsf201.

LOCUS XELRGASBC 375 bp ds-DNA VRT 05-JUN-1991
DEFINITION x.borealis somatic 5s rrna gene, clone pxbsf201.
ACCESSION K01537
KEYWORDS 5S ribosomal RNA; ribosomal RNA.
SOURCE xenopus borealis dna, clone pxbsf201.
ORGANISM Xenopus laevis
Eukaryota; Animalia; Chordata; Vertebrata; Amphibia; Lissamphibia;
Anura; Archeobatrachia; Pipoidae; Pipidae; Xenopodinae.
REFERENCE 1 (bases 1 to 375)
AUTHORS Razvi,F., Gargiulo,G. and Worcel,A.
TITLE a simple procedure for parallel sequence analysis of both strands
of 5'-labeled dna
JOURNAL Gene 23, 175-183 (1983)
STANDARD full automatic
COMMENT NCBI gi: 214699
FEATURES Location/Qualifiers
source 1..375
/organism="Xenopus laevis"
misc_feature complement(1..29)
/note="putative VECTOR sequence Vector pUC19 (M11662);
putative"
rRNA 80..199
/note="5s rrna"
misc_feature 286..375
/note="putative VECTOR sequence Bacteriophage M13mp18
(M11454); putative"
BASE COUNT 80 a 116 c 96 g 83 t
ORIGIN 2 bp upstream of alui site.

Initial Score = 17 Optimized Score = 20 Significance = 7.51
Residue Identity = 95% Matches = 20 Mismatches = 1
Gaps = 0 Conservative Substitutions = 0

CATACCACCCTGAAAGTGCCCGATATCGTCTGATCTCGGAAGCCAAGCAGGGTCGGGCCTGGTTAGTACTTG
90 100 110 120 130 140 150 160

X 10 X
GTCCTAGGCTTTTGCACTTTT
||| |||||
GATGGGAGACCGCCTGGGAATACCAGGTGTCGTAGGCTTTTGCACTTTTGCCATTCTGAGTAACAGCAGGGG
170 180 190 200 210 220 230
GCAGTCTCCTCCATGCATTTTTCTTTCCCCGAACAGCCGGATCCCCGGGAATTCAGTGGCCGTCGTTTTACA
240 250 260 270 280 290 300
ACGTC

```
> 0 (
01 10 IntelliGenetics
> 0 (
```

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file jan11b.res made by on Wed 11 Jan 95 12:32:44-PST.

```
Query sequence being compared:      CL17 (1-38)
Number of sequences searched:      302507
Number of scores above cutoff:      4183
```

Results of the initial comparison of CL17 (1-38) with:

- Data bank : EMBL-NEW 10, all entries
- Data bank : GenBank 85, all entries
- Data bank : GenBank-NEW 10, all entries
- Data bank : HIV-NA 7, all entries
- Data bank : Issued_NA , all entries
- Data bank : N-GeneSeq 16.3, all entries
- Data bank : UEMBL 40_85, all entries
- Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	2		
Randomization group	0		
Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	9	10	4.78

Times:	CPU	Total Elapsed
	00:13:23.07	00:13:42.00

```
Number of residues:      276734581
Number of sequences searched: 302507
Number of scores above cutoff: 4183
```

Cut-off raised to 4.
Cut-off raised to 5.
Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 9.
Cut-off raised to 11.
Cut-off raised to 13.
Cut-off raised to 14.

Cut-off raised to 15.
 Cut-off raised to 16.
 Cut-off raised to 17.
 Cut-off raised to 18.
 Cut-off raised to 19.
 Cut-off raised to 20.
 Cut-off raised to 21.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame

**** 3 standard deviations above mean ****						
1. HSU09850	Human zinc finger protein (ZN	3908	28	30	3.98	0
2. USMURBS1A	Ustilago maydis URBS1 protein	3987	28	31	3.98	0
3. S76114	{right virus-host integration	569	27	29	3.77	0
4. OCACE3P	O.cuniculus DNA for angiotens	978	27	28	3.77	0
5. HUMCOUPII	Homo sapiens chick ovalbumin	2268	27	28	3.77	0
6. HSCOUPII	Homo sapiens chick ovalbumin	2268	27	28	3.77	0
7. PALHISH2H3	P.lividus histone H3 and H2A	2291	27	29	3.77	0
8. RABACEA	Rabbit angiotensin converting	2409	27	28	3.77	0
9. MMGK5	Mouse glandular kallikrein ge	3610	27	30	3.77	0
10. LUMHBC	Earthworm (L.terrestris) extr	4037	27	31	3.77	0
11. MMIFOR	M.musculus mRNA for formin (i	4241	27	30	3.77	0
12. OCANCOE	O.cuniculus mRNA for angioten	4800	27	28	3.77	0
13. MMLDF	M.musculus limb deformity mRN	4973	27	30	3.77	0
14. CELB0280	Caenorhabditis elegans cosmid	41088	27	28	3.77	0
15. CEB0280	Caenorhabditis elegans cosmid	41088	27	28	3.77	0
16. RICR20321A	Rice cDNA, partial sequence (271	26	29	3.56	0
17. T21884	3892 Arabidopsis thaliana cDN	278	26	26	3.56	0
18. RIC1140A	Rice cDNA, partial sequence (353	26	28	3.56	0
19. T09049	EST06941 Homo sapiens cDNA cl	394	26	27	3.56	0
20. RATMLCB1	Rat cardiac myosin light chai	549	26	28	3.56	0
21. HUMITILC03	Human inter-alpha-trypsin inh	618	26	29	3.56	0
22. HUMMHDVB2	Human MHC class II HLA-DV-bet	745	26	29	3.56	0
23. MUSNAKATPQ	Mouse Na,K-ATPase beta2 subun	1128	26	29	3.56	0
24. PSELINC	P.paucimobilis linC gene for	1148	26	26	3.56	0
25. PPLINC	P.paucimobilis linC gene for	1148	26	26	3.56	0
26. BOVPROA	Bovine protamine gene P1 alle	1340	26	27	3.56	0
27. BOVPROB	Bovine protamine gene P1 alle	1369	26	27	3.56	0
28. HUMMHDQBAA	Human MHC class II HLA-DQB3 p	1416	26	29	3.56	0
29. MUSIGHYC1	Mouse Ig heavy-chain variable	1599	26	30	3.56	0
30. MMIGVH28	Mouse immunoglobulin J558 V(H	1599	26	30	3.56	0

Query sequence being compared: CL17 (1-38)
 Number of sequences optimized: 4183

Results of the optimized comparison of CL17 (1-38) with:

Data bank : EMBL-NEW 10, all entries
 Data bank : GenBank 85, all entries

Data bank : GenBank-NEW 10, all entries
 Data bank : HIV-NA 7, all entries
 Data bank : Issued_NA , all entries
 Data bank : N-GeneSeq 16.3, all entries
 Data bank : UEMBL 40_85, all entries
 Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	2		
Randomization group	0		

Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	24	26	1.45

Times:	CPU	Total Elapsed
	00:00:49.98	00:01:03.00

Number of residues:	16505319
Number of sequences optimized:	4183

The scores below are sorted by optimized score.
 Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame

**** 4 standard deviations above mean ****						
1. SHPIGFIIA	Ovis aries insulin-like growt	1036	25	31	4.82	0
2. USMURBS1A	Ustilago maydis URBS1 protein	3987	28	31	4.82	0
3. Q61404	Human brain Expressed Sequenc	361	23	31	4.82	0
4. M79245	EST01393 Homo sapiens cDNA cl	361	23	31	4.82	0
5. HSCD19	H.sapiens RNA for CD19.	1910	21	31	4.82	0
6. HUMCD19W01	Human CD19 gene, exons 1-4.	1916	21	31	4.82	0
7. N90612	CD19 cDNA.	1921	21	31	4.82	0
8. Q21172	Human CD19 antigen coding seq	1922	21	31	4.82	0
9. LUMHBC	Earthworm (L.terrestris) extr	4037	27	31	4.82	0
10. OANIGFII4	Ovis aries insulin-like growt	547	25	31	4.82	0
11. HUMCD19A	Human CD19 gene, complete cds	8743	21	31	4.82	0
12. HUMANTCD	Human differentiation antigen	1922	21	31	4.82	0
13. HUMCSPC	Human cell surface protein CD	2096	21	31	4.82	0

14.	OAINIGFII4	Ovis aries insulin-like growt	547	25	31	4.82	0
15.	MMLDF	M.musculus limb deformity mRN	4973	27	30	4.13	0
16.	MMGK5	Mouse glandular kallikrein ge	3610	27	30	4.13	0
17.	MMIFOR	M.musculus mRNA for formin (i	4241	27	30	4.13	0
18.	HSU09850	Human zinc finger protein (ZN	3908	28	30	4.13	0
19.	US-07-841-646-3	Sequence 3, Application US	17410	21	30	4.13	0
20.	HSTUBAG	Human gene for alpha-tubulin	4087	24	30	4.13	0
21.	RNA1B	Rat mRNA for the alpha-1B adr	2086	21	30	4.13	0
22.	RATRDLX	Rat homeoprotein (rDlx) mRNA,	1390	21	30	4.13	0
23.	DMLABR	Drosophila melanogaster F24 m	2159	25	30	4.13	0
24.	Q53142	Sequence encoding osteogenic	17410	21	30	4.13	0
25.	MMHOX4	M.musculus genes HOX-4.4 and	8601	24	30	4.13	0
26.	DMLABG1	Drosophila melanogaster F24 l	1846	25	30	4.13	0
27.	RATGENOME	Rat gene for alpha 1B adrener	2387	21	30	4.13	0
28.	OSRGP1	Rice rgp1 mRNA for a ras-rela	1303	22	30	4.13	0
29.	MUSIGHYC1	Mouse Ig heavy-chain variable	1599	26	30	4.13	0
30.	MMIGVH28	Mouse immunoglobulin J558 V(H	1599	26	30	4.13	0

1. CL17 (1-38)

SHPIGFIIA Ovis aries insulin-like growth factor II (IGF-II)

LOCUS	SHPIGFIIA	1036 bp ss-mRNA	MAM	22-JUL-1993
DEFINITION	Ovis aries insulin-like growth factor II (IGF-II) mRNA, complete cds.			
ACCESSION	M89788			
KEYWORDS	insulin-like growth factor II.			
SOURCE	Ovis aries (strain Coopworth) (library: random primed cDNA) lamb liver cDNA to mRNA.			
ORGANISM	Ovis aries Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae.			
REFERENCE	1 (bases 1 to 1036)			
AUTHORS	Demmer, J., Hill, D.F. and Petersen, G.B.			
TITLE	Characterization of two sheep insulin-like growth factor II cDNAs with different 5'-untranslated regions			
JOURNAL	Biochim. Biophys. Acta 1173, 79-80 (1993)			
STANDARD	full automatic			
COMMENT	NCBI gi: 165940			
FEATURES	Location/Qualifiers			
source	1..1036 /organism="Ovis aries" /strain="Coopworth" /dev_stage="lamb" /sequenced_mol="cDNA to mRNA" /tissue_type="liver" /tissue_lib="random primed cDNA"			
sig_peptide	102..173 /gene="IGF-II" /codon_start=1			
CDS	102..641 /gene="IGF-II" /note="NCBI gi: 165941" /codon_start=1 /product="insulin-like growth factor II" /translation="MGITAGKSMALLAFLAFASCCYAAYRPSETLCGGELVDTLQFVCGDRGFYFSRPSSRINRRSRGIVECCFRSCDLALLETYCAAPAKSERDVSASTTVLP			

```

mat_peptide      DDFTAYPVGKFFQSDTWKQSTQRLRRGLPAFLRARRGRTLAKLEALREAKSHRPLIA
                  LPTQDPATHGGASSEASSD"
                  174..374
                  /gene="IGF-II"
                  /codon_start=1
                  /product="insulin-like growth factor II"

```

```

BASE COUNT      220 a    368 c    236 g    212 t
ORIGIN

```

```

Initial Score    =    25  Optimized Score =    31  Significance =  4.82
Residue Identity =   80%  Matches          =    33  Mismatches   =    5
Gaps             =     3  Conservative Substitutions =    0

```

```

                  X          10          20          30          X
                  TC--GACTCCTCTTCCTCCTCCACCTCCTCCTCC-CATGCA
                  ||  ||||| ||||| ||||| ||||| ||| | | ||
GGTAGCTTCTCCTCGGAGGCAGCCTTCCAGACTCCTCCTCCTCCTCCTCCTCATCCTCCTTCAGCCCCA
          10          20      X  30          40          50          60      X  70

GCGAGCCTCCTGTCCAGCTGCAGACATCAATGGGGATCACAGCAGGAAAGTCGATGCTGGCGCTTCTTGCCT
          80          90          100          110          120          130          140

TCTTGGCCTTCGCCTCGTGCTG
          150          160

```

> 0 (<
OI IO IntelliGenetics
> 0 (<

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file jan11c.res made by on Wed 11 Jan 95 12:47:51-PST.

Query sequence being compared: CL26 (1-21)
Number of sequences searched: 302507
Number of scores above cutoff: 4620

Results of the initial comparison of CL26 (1-21) with:
Data bank : EMBL-NEW 10, all entries
Data bank : GenBank 85, all entries
Data bank : GenBank-NEW 10, all entries
Data bank : HIV-NA 7, all entries
Data bank : Issued_NA , all entries
Data bank : N-GeneSeq 16.3, all entries
Data bank : UEMBL 40_85, all entries
Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		
Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	6	7	2.92

Times:	CPU	Total Elapsed
	00:12:46.03	00:12:48.00

Number of residues: 276734581
Number of sequences searched: 302507
Number of scores above cutoff: 4620

Cut-off raised to 4.
Cut-off raised to 5.
Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.
Cut-off raised to 10.
Cut-off raised to 11.

Cut-off raised to 12.
 Cut-off raised to 13.
 Cut-off raised to 14.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 3 standard deviations above mean ****						
1. XELRGASBC	x.borealis somatic 5s rrna ge	375	17	20	3.77	0
2. XELRGASBA	X.borealis somatic 5S rRNA ge	858	17	20	3.77	0
3. XBRNA2	Xenopus borealis gene for 5S	858	17	20	3.77	0
4. RATMAP1A	Rat MAP-1 gene encoding major	1402	17	17	3.77	0
5. RATTKG1	Rat T-kininogen (T-KG) gene,	1903	17	17	3.77	0
6. RNAMDX23	R.norvegicus S-adenosylmethio	2021	17	17	3.77	0
7. ZDHRGP	Z.diploperennis gene for hydr	4478	17	17	3.77	0
8. MMT1CPS	Mouse Tla region Tlc pseudoge	8147	17	17	3.77	0
9. RATSADMEDC	Rat AdoMetDC gene, complete C	17167	17	17	3.77	0
10. CELF28F5	Caenorhabditis elegans cosmid	32903	17	17	3.77	0
11. CELF28F5	Caenorhabditis elegans cosmid	32903	17	17	3.77	0
12. CHNTXX	Tobacco chloroplast genome DN	155844	17	17	3.77	0
13. N60861	Fragment of plasmid PXC204 en	146	16	16	3.43	0
14. T24747	EST322 Homo sapiens cDNA clon	186	16	16	3.43	0
15. HS7476	EST322 Homo sapiens cDNA clon	186	16	16	3.43	0
16. NVIRGAA	Newt (Notophthalmus viridesce	235	16	18	3.43	0
17. NV5SRRN	Notophthalmus viridescens 5S	235	16	18	3.43	0
18. N60862	Fragment of plasmid PXC204 en	288	16	16	3.43	0
19. XELCRLB	Xenopus laevis caerulein prec	301	16	16	3.43	0
20. PABKTANT	BK virus 5' end of early regi	332	16	16	3.43	0
21. XLCAER1	Xenopus laevis mRNA fragment	370	16	16	3.43	0
22. XELCRLA	Xenopus laevis caerulein prec	370	16	16	3.43	0
23. T08475	EST06366 Homo sapiens cDNA cl	383	16	16	3.43	0
24. XELCRLG35	Xenopus laevis caerulein type	391	16	16	3.43	0
25. XELCRLI	X.laevis caerulein mRNA, clon	395	16	16	3.43	0
26. N60858	Sequence of plasmid PXC102 en	397	16	16	3.43	0
27. PVBRESWW	Human papovavirus BK (strain	426	16	16	3.43	0
28. HUMUT2361	Human STS UT2361.	446	16	16	3.43	0
29. BRRRPL37A	Brassica rapa ribosomal prote	446	16	16	3.43	0
30. N50145	Sequence of enhancer DNA segm	451	16	16	3.43	0

Query sequence being compared: CL26 (1-21)
 Number of sequences optimized: 4620

Results of the optimized comparison of CL26 (1-21) with:

Data bank : EMBL-NEW 10, all entries
 Data bank : GenBank 85, all entries
 Data bank : GenBank-NEW 10, all entries
 Data bank : HIV-NA 7, all entries
 Data bank : Issued_NA , all entries
 Data bank : N-GeneSeq 16.3, all entries

Data bank : UEMBL 40_85, all entries
Data bank : VectorBank 9, all entries

PCR/US44/11646

08/324001

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		
Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	14	15	0.80
Times:	CPU	Total Elapsed	
	00:01:01.97	00:01:07.00	
Number of residues:	24455601		
Number of sequences optimized:	4620		

The scores below are sorted by optimized score.
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 7 standard deviations above mean ****					
1. XELRGASBC	x.borealis somatic 5s rrna ge	375	17	20	7.51	0
2. XELRGASBA	X.borealis somatic 5S rRNA ge	858	17	20	7.51	0
3. XBRNA2	Xenopus borealis gene for 5S	858	17	20	7.51	0
	**** 6 standard deviations above mean ****					
4. XELRGASL	X.laevis somatic 5S rRNA gene	888	16	19	6.26	0
	**** 5 standard deviations above mean ****					
5. NV5SRRN	Notophthalmus viridescens 5S	235	16	18	5.01	0
6. NVIRGAA	Newt (Notophthalmus viridescence)	235	16	18	5.01	0
7. XELRGADB	x.borealis oocyte 5s dna.	761	15	18	5.01	0
8. XBRNA1	Xenopus borealis genes (three	761	15	18	5.01	0
	**** 3 standard deviations above mean ****					
9. MMT1CPS	Mouse Tla region Tlc pseudoge	8147	17	17	3.76	0
10. ZDHRGP	Z.diploperennis gene for hydr	4478	17	17	3.76	0
11. RATTKG1	Rat T-kininogen (T-KG) gene,	1903	17	17	3.76	0
12. CELF28F5	Caenorhabditis elegans cosmid	32903	17	17	3.76	0

PCT/WS94/11616

13. RNAMDX23	R.norvegicus S-adenosylmethio	2021	17	17	3.76	0
14. CELF28F5	Caenorhabditis elegans cosmid	32903	17	17	3.76	0

15. RATMAP1A	Rat MAP-1 gene encoding major	1402	17	17	3.76	0
16. CHNTXX	Tobacco chloroplast genome DN	155844	17	17	3.76	0
17. RATSADMEDC	Rat AdoMetDC gene, complete C	17167	17	17	3.76	0
18. ONHGHCOHO	Oncorhynchus kisutch (coho sa	1201	15	17	3.76	0
19. STREIPEPA	Streptococcus salivarius phos	2259	15	17	3.76	0
20. XLXK70A	Xenopus laevis XK70A gene for	6266	15	17	3.76	0
21. CEF54C8	Caenorhabditis elegans cosmid	23000	15	17	3.76	0
22. CEF54C8	Caenorhabditis elegans cosmid	23000	15	17	3.76	0
23. MIDACYTB	O.aries mitochondrion cytb ge	1140	13	17	3.76	0
24. SV4EV211	SV40 variant genome ev-2114,	100	14	17	3.76	0
25. DR07DC14Z	Drosophila melanogaster (subc	3625	13	17	3.76	0
26. T10577	hbc220 Homo sapiens cDNA clon	560	13	17	3.76	0
27. NEUFRG	Neurospora crassa mRNA sequen	4631	13	17	3.76	0
28. HSA26A071	H. sapiens partial cDNA seque	347	13	17	3.76	0
29. HSA39H101	H. sapiens partial cDNA seque	345	12	17	3.76	0
30. PCT-US93-04648-1	Sequence 15, Application	10596	14	17	3.76	0

1. CL26 (1-21)

XELRGASBC x.borealis somatic 5s rrna gene, clone pxbsf201.

LOCUS XELRGASBC 375 bp ds-DNA VRT 05-JUN-1991

DEFINITION x.borealis somatic 5s rrna gene, clone pxbsf201.

ACCESSION K01537

KEYWORDS 5S ribosomal RNA; ribosomal RNA.

SOURCE xenopus borealis dna, clone pxbsf201.

ORGANISM Xenopus laevis
Eukaryota; Animalia; Chordata; Vertebrata; Amphibia; Lissamphibia;
Anura; Archeobatrachia; Pipoidae; Pipidae; Xenopodinae.

REFERENCE 1 (bases 1 to 375)

AUTHORS Razvi,F., Gargiulo,G. and Worcel,A.

TITLE a simple procedure for parallel sequence analysis of both strands
of 5'-labeled dna

JOURNAL Gene 23, 175-183 (1983)

STANDARD full automatic

COMMENT NCBI gi: 214699

FEATURES Location/Qualifiers

source 1..375
/organism="Xenopus laevis"

misc_feature complement(1..29)
/note="putative VECTOR sequence Vector pUC19 (M11662);
putative"

rRNA 80..199
/note="5s rrna"

misc_feature 286..375
/note="putative VECTOR sequence Bacteriophage M13mp18
(M11454); putative"

BASE COUNT 80 a 116 c 96 g 83 t

ORIGIN 2 bp upstream of alui site.

Initial Score = 17 Optimized Score = 20 Significance = 7.51

Residue Identity = 95% Matches = 20 Mismatches = 1

Gaps = 0 Conservative Substitutions = 0

CATACCACCCTGAAAGTGCCCGATATCGTCTGATCTCGGAAGCCAAGCAGGGTCGGGCCTGGTTAGTACTTG

90 100 110 120 130 140 150 160

X 10 X
GTCCTAGGCTTTTGCACCTTT
III IIIIIIIIIIIIIIIIIII
GATGGGAGACCGCCTGGGAATACCAGGTGTCGTAGGCTTTTGCACCTTTGCCATTCTGAGTAACAGCAGGGG
170 180 190 200 210 220 230
GCAGTCTCCTCCATGCATTTTCTTTCCCCGAACAGCCGGATCCCCGGGAATCACTGGCCGTCGTTTTACA
240 250 260 270 280 290 300
ACGTC

> 0 (<
01 10 IntelliGenetics
> 0 (<

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file jan11d.res made by on Wed 11 Jan 95 13:02:40-PST.

Query sequence being compared: CL16' (1-21)
Number of sequences searched: 302507
Number of scores above cutoff: 4881

Results of the initial comparison of CL16' (1-21) with:

Data bank : EMBL-NEW 10, all entries
Data bank : GenBank 85, all entries
Data bank : GenBank-NEW 10, all entries
Data bank : HIV-NA 7, all entries
Data bank : Issued_NA , all entries
Data bank : N-GeneSeq 16.3, all entries
Data bank : UEMBL 40_85, all entries
Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		
Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	6	7	3.01

Times:	CPU	Total Elapsed
	00:13:19.97	00:13:35.00

Number of residues: 276734581
Number of sequences searched: 302507
Number of scores above cutoff: 4881

Cut-off raised to 4.
Cut-off raised to 5.
Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.
Cut-off raised to 10.
Cut-off raised to 11.
Cut-off raised to 12.

Cut-off raised to 13.

Cut-off raised to 14.

The scores below are sorted by initial score.

Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Init. Opt. Score	Sig.	Frame
**** 3 standard deviations above mean ****						
1. BTRPTDNAE	B.taurus repeat region DNA.	482	18	18	3.99	0
2. RABTCRGAM	Rabbit T-cell receptor gamma	147	17	17	3.65	0
3. Q77574	Human genome fragment. (Prefe	200	17	17	3.65	0
4. HSAAACMHG	H. sapiens putatively transcr	200	17	17	3.65	0
5. ATTS1638	A. thaliana transcribed seque	274	17	17	3.65	0
6. TBILTA124	T.brucei mRNA for variant sur	1688	17	17	3.65	0
7. TBRVSG	T.brucei rhodensiense mRNA fo	1732	17	17	3.65	0
8. U01312	Streptococcus pyogenes JRS4 p	1823	17	17	3.65	0
9. S52562	LH-2=LIM/homeodomain protein	2072	17	17	3.65	0
10. HUMSWX167	Human chromosome X STS sWXD16	239	16	17	3.32	0
11. GCREG35	Galago Alu repeat type II, GA	245	16	16	3.32	0
12. HSA127WB5	H. sapiens (D1S505) DNA segme	319	16	16	3.32	0
13. NEUMTOLI2	N. crassa mitochondrial oli2	335	16	16	3.32	0
14. PLYORIA	Human polyomavirus BK (strain	375	16	16	3.32	0
15. M88810	CEL01E12 Caenorhabditis elega	394	16	16	3.32	0
16. S67523	early gene, late gene (contro	401	16	16	3.32	0
17. ATTS2283	A. thaliana transcribed seque	408	16	16	3.32	0
18. HS311VF9	H. sapiens (D5S662) DNA segme	414	16	16	3.32	0
19. PLYORIB	Human polyomavirus BK (strain	424	16	16	3.32	0
20. ATTS1882	A. thaliana transcribed seque	429	16	16	3.32	0
21. PVBECR522	Human papovavirus BK, Gardner	455	16	16	3.32	0
22. SYNECR530	BKV hybrid (tr-530) early tra	487	16	16	3.32	0
23. SYNECR532	BKV hybrid (tr-532) early tra	515	16	16	3.32	0
24. SYNECR531	BKV hybrid (tr-531) early tra	558	16	16	3.32	0
25. PVBECR501	Human papovavirus BK, Gardner	559	16	16	3.32	0
26. CEZMTTGP	Green turtle mitochondrion tr	620	16	16	3.32	0
27. Q58456	BK enhancer-adenovirus-2 late	642	16	16	3.32	0
28. Q54210	BK enhancer-adenovirus 2 late	642	16	16	3.32	0
29. HUMRPO	Human gene for ret proto-onco	678	16	16	3.32	0
30. ZEFTRANB	Danio rerio mRNA, Tci-like tr	706	16	16	3.32	0

Query sequence being compared: CL16' (1-21)

Number of sequences optimized: 4881

Results of the optimized comparison of CL16' (1-21) with:

Data bank : EMBL-NEW 10, all entries

Data bank : GenBank 85, all entries

Data bank : GenBank-NEW 10, all entries

Data bank : HIV-NA 7, all entries

Data bank : Issued_NA , all entries

Data bank : N-GeneSeq 16.3, all entries

Data bank : UEMBL 40_85, all entries

Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		
Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	14	15	0.78
Times:	CPU	Total Elapsed	
	00:01:03.97	00:01:08.00	
Number of residues:	23291943		
Number of sequences optimized:	4881		

The scores below are sorted by optimized score.
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. BTRPTDNAE	B.taurus repeat region DNA.	482	18	18	5.15	0
2. MUSMA	Mouse mRNA for ORF.	7222	16	18	5.15	0
3. S92205	rna12+=pre-rRNA maturation IS	3587	15	18	5.15	0
4. ZEFTRAN	Danio rerio Tc1-like transpos	1205	16	18	5.15	0
**** 3 standard deviations above mean ****						
5. HSAAACMHG	H. sapiens putatively transcr	200	17	17	3.86	0
6. TBILTA124	T.brucei mRNA for variant sur	1688	17	17	3.86	0
7. ATTS1638	A. thaliana transcribed seque	274	17	17	3.86	0
8. TBRVSG	T.brucei rhodensiense mRNA fo	1732	17	17	3.86	0
9. S52562	LH-2=LIM/homeodomain protein	2072	17	17	3.86	0
10. HUMSWX167	Human chromosome X STS sWXD16	239	16	17	3.86	0
11. Q77574	Human genome fragment. (Prefe	200	17	17	3.86	0
12. U01312	Streptococcus pyogenes JRS4 p	1823	17	17	3.86	0
13. RABTCRGAM	Rabbit T-cell receptor gamma	147	17	17	3.86	0
14. T16193	IB3700 Homo sapiens cDNA 3'en	498	15	17	3.86	0
15. ZEFTRAN	Danio rerio Tc1-like transpos	1241	15	17	3.86	0
16. SSIS1139	S.salivarius insertion sequen	1717	15	17	3.86	0
17. YSKSTE12X	Kluyveromyces lactis STE12 ge	2678	15	17	3.86	0

18. CEZC84	Caenorhabditis elegans cosmid	38955	15	17	3.86	0
19. CEZC84	Caenorhabditis elegans cosmid	38955	15	17	3.86	0
20. CEZC84	Caenorhabditis elegans cosmid	38955	15	17	3.86	0
21. M28728	Figure 1. (B) Sequences in wt	51	14	17	3.86	0
22. Q38699	Oligonucleotide 7 to insert g	63	14	17	3.86	0
23. SV4MNKR5	simian virus 40/african green	115	14	17	3.86	0
24. HSBA7H052	H. sapiens partial cDNA seque	231	14	17	3.86	0
25. SV4MNKR4	simian virus 40/african green	250	14	17	3.86	0
26. SV4STA	Rhesus macaque polyoma virus	384	14	17	3.86	0
27. SV4MNKR3	simian virus 40/african green	593	14	17	3.86	0
28. SV4STA4	Rhesus macaque polyoma virus	694	14	17	3.86	0
29. HUMRAB6A	Homo sapiens GTP-binding prot	740	14	17	3.86	0
30. HSRAB6A	Homo sapiens GTP-binding prot	740	14	17	3.86	0

1. CL16' (1-21)

BTRPTDNAE B.taurus repeat region DNA.

LOCUS	BTRPTDNAE	482 bp	DNA	MAM	16-AUG-1993
DEFINITION	B.taurus repeat region DNA.				
ACCESSION	Z25529				
KEYWORDS	repeat region.				
SOURCE	cattle.				
ORGANISM	Bos taurus				
	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;				
	Theria; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae.				
REFERENCE	1 (bases 1 to 482)				
AUTHORS	Szemraj,J., Plucienniczak,G., Jaworski,J. and Plucienniczak,A.				
TITLE	Evidence for homological recombination with participation of the				
	bovine alu-like sequences				
JOURNAL	Unpublished				
STANDARD	full automatic				
REFERENCE	2 (bases 1 to 482)				
AUTHORS	Plucienniczak,A.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-AUG-1993) PLUCIENNICZAK A., PP TERPOL, LABORATORY OF				
	GENETIC ENGINEERING, P.O.W. 57, SIERADZ, POLAND, 98-200				
STANDARD	full automatic				
COMMENT	NCBI gi: 396758				
FEATURES	Location/Qualifiers				
source	1..482				
	/organism="Bos taurus"				
	/clone="pUJ3.24"				
	/dev_stage="calf"				
	/tissue_type="thymus"				
repeat_unit	133..482				
	/partial				
	/note="Truncated 5' part of BDDF."				
	/rpt_type=DISPERSED				
	/evidence=experimental				
	/rpt_family="Bovine Dimer Driven Family (BDDF)"				
	/label=BDDF				
	/citation=[1]				
repeat_unit	373..426				
	/partial				
	/note="5' part of bovine alu-like monomer."				
	/rpt_type=FLANKING				

/evidence=experimental
/rpt_family="bovine alu-like"
/citation=[1]

BASE COUNT 135 a 109 c 124 g 114 t
ORIGIN

Initial Score = 18 Optimized Score = 18 Significance = 5.15
Residue Identity = 85% Matches = 18 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

GGGTCGATGGTGGAGAGGTCGTGACGAGAATGTAGTCCACTGGAGAGGGGAATGGCAAACTACTTCAGTATT
 340 350 360 370 380 390 400

 X 10 20
 AAAAGTGCAAAAGCCTAGGAC
 ||||| ||||| |||||

CTTGCCTTGAGAACCCCATGAACGTATGAAAAGGGCAAAAGCATAGGATAGCTGAAAGAGGAACTCCCCAGT
 410 420 430 X 440 450 X 460 470

CGATAGG
 480

> 0 (
01 10 IntelliGenetics
> 0 (

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file jan11f.res made by on Wed 11 Jan 95 12:47:30-PST.

Query sequence being compared: CL26' (1-21)
Number of sequences searched: 302507
Number of scores above cutoff: 4881

Results of the initial comparison of CL26' (1-21) with:
Data bank : EMBL-NEW 10, all entries
Data bank : GenBank 85, all entries
Data bank : GenBank-NEW 10, all entries
Data bank : HIV-NA 7, all entries
Data bank : Issued_NA , all entries
Data bank : N-GeneSeq 16.3, all entries
Data bank : UEMBL 40_85, all entries
Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix Unitary K-tuple 4
Mismatch penalty 1 Joining penalty 30
Gap penalty 1.00 Window size 14
Gap size penalty 0.33
Cutoff score 1
Randomization group 0
Initial scores to save 30 Alignments to save 30
Optimized scores to save 30 Display context 100

SEARCH STATISTICS

Scores: Mean Median Standard Deviation
6 7 3.01

Times: CPU Total Elapsed
00:13:03.06 00:13:28.00

Number of residues: 276734581
Number of sequences searched: 302507
Number of scores above cutoff: 4881

Cut-off raised to 4.
Cut-off raised to 5.
Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.
Cut-off raised to 10.
Cut-off raised to 11.

Cut-off raised to 12.
 Cut-off raised to 13.
 Cut-off raised to 14.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 3 standard deviations above mean ****						
1. BTRPTDNAE	B.taurus repeat region DNA.	482	18	18	3.99	0
2. RABTCRGAM	Rabbit T-cell receptor gamma	147	17	17	3.65	0
3. Q77574	Human genome fragment. (Prefe	200	17	17	3.65	0
4. HSAAACMHG	H. sapiens putatively transcr	200	17	17	3.65	0
5. ATTS1638	A. thaliana transcribed seque	274	17	17	3.65	0
6. TBILTA124	T.brucei mRNA for variant sur	1688	17	17	3.65	0
7. TBRVSG	T.brucei rhodensiense mRNA fo	1732	17	17	3.65	0
8. U01312	Streptococcus pyogenes JRS4 p	1823	17	17	3.65	0
9. S52562	LH-2=LIM/homeodomain protein	2072	17	17	3.65	0
10. HUMSWX167	Human chromosome X STS sWxD16	239	16	17	3.32	0
11. GCREG35	Galago Alu repeat type II, GA	245	16	16	3.32	0
12. HSA127WB5	H. sapiens (D1S505) DNA segme	319	16	16	3.32	0
13. NEUMTOLI2	N. crassa mitochondrial oli2	335	16	16	3.32	0
14. PLYORIA	Human polyomavirus BK (strain	375	16	16	3.32	0
15. M88810	CEL01E12 Caenorhabditis elega	394	16	16	3.32	0
16. S67523	early gene, late gene (contro	401	16	16	3.32	0
17. ATTS2283	A. thaliana transcribed seque	408	16	16	3.32	0
18. HS311VF9	H. sapiens (D5S662) DNA segme	414	16	16	3.32	0
19. PLYORIB	Human polyomavirus BK (strain	424	16	16	3.32	0
20. ATTS1882	A. thaliana transcribed seque	429	16	16	3.32	0
21. PVBECR522	Human papovavirus BK, Gardner	455	16	16	3.32	0
22. SYNECR530	BKV hybrid (tr-530) early tra	487	16	16	3.32	0
23. SYNECR532	BKV hybrid (tr-532) early tra	515	16	16	3.32	0
24. SYNECR531	BKV hybrid (tr-531) early tra	558	16	16	3.32	0
25. PVBECR501	Human papovavirus BK, Gardner	559	16	16	3.32	0
26. CEZMTTGP	Green turtle mitochondrion tr	620	16	16	3.32	0
27. Q58456	BK enhancer-adenovirus-2 late	642	16	16	3.32	0
28. Q54210	BK enhancer-adenovirus 2 late	642	16	16	3.32	0
29. HUMRPO	Human gene for ret proto-onco	678	16	16	3.32	0
30. ZEFTRANB	Danio rerio mRNA, Tc1-like tr	706	16	16	3.32	0

Query sequence being compared: CL26' (1-21)
 Number of sequences optimized: 4881

Results of the optimized comparison of CL26' (1-21) with:
 Data bank : EMBL-NEW 10, all entries
 Data bank : GenBank 85, all entries
 Data bank : GenBank-NEW 10, all entries
 Data bank : HIV-NA 7, all entries
 Data bank : Issued_NA , all entries
 Data bank : N-GeneSeq 16.3, all entries

Data bank : UEMBL 40_85, all entries
 Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		
Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	14	15	0.78
Times:	CPU	Total Elapsed	
	00:01:01.91	00:01:09.00	
Number of residues:	23291943		
Number of sequences optimized:	4881		

The scores below are sorted by optimized score.
 Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 5 standard deviations above mean ****					
1. BTRPTDNAE	B.taurus repeat region DNA.	482	18	18	5.15	0
2. MUSMA	Mouse mRNA for ORF.	7222	16	18	5.15	0
3. S92205	rna12+=pre-rRNA maturation [S	3587	15	18	5.15	0
4. ZEFTRAN	Danio rerio Tc1-like transpos	1205	16	18	5.15	0
	**** 3 standard deviations above mean ****					
5. HSAAACMHG	H. sapiens putatively transcr	200	17	17	3.86	0
6. TBILTA124	T.brucei mRNA for variant sur	1688	17	17	3.86	0
7. ATTS1638	A. thaliana transcribed seque	274	17	17	3.86	0
8. TBRVSG	T.brucei rhodensiense mRNA fo	1732	17	17	3.86	0
9. S52562	LH-2=LIM/homeodomain protein	2072	17	17	3.86	0
10. HUMSWX167	Human chromosome X STS sWXD16	239	16	17	3.86	0
11. Q77574	Human genome fragment. (Prefe	200	17	17	3.86	0
12. U01312	Streptococcus pyogenes JRS4 p	1823	17	17	3.86	0
13. RABTCRGAM	Rabbit T-cell receptor gamma	147	17	17	3.86	0
14. T16193	IB3700 Homo sapiens cDNA 3'en	498	15	17	3.86	0
15. ZEFTRAN	Danio rerio Tc1-like transpos	1241	15	17	3.86	0
16. SSIS1139	S.salivarius insertion sequen	1717	15	17	3.86	0

17. YSKSTE12X	Kluyveromyces lactis STE12 ge	2678	15	17	3.86	0
18. CEZC84	Caenorhabditis elegans cosmid	38955	15	17	3.86	0
19. CEZC84	Caenorhabditis elegans cosmid	38955	15	17	3.86	0
20. CEZC84	Caenorhabditis elegans cosmid	38955	15	17	3.86	0
21. M28728	Figure 1. (B) Sequences in wt	51	14	17	3.86	0
22. Q38699	Oligonucleotide 7 to insert g	63	14	17	3.86	0
23. SV4MNKR5	simian virus 40/african green	115	14	17	3.86	0
24. HSBA7H052	H. sapiens partial cDNA seque	231	14	17	3.86	0
25. SV4MNKR4	simian virus 40/african green	250	14	17	3.86	0
26. SV4STA	Rhesus macaque polyoma virus	384	14	17	3.86	0
27. SV4MNKR3	simian virus 40/african green	593	14	17	3.86	0
28. SV4STA4	Rhesus macaque polyoma virus	694	14	17	3.86	0
29. HUMRAB6A	Homo sapiens GTP-binding prot	740	14	17	3.86	0
30. HSRAB6A	Homo sapiens GTP-binding prot	740	14	17	3.86	0

1. CL26' (1-21)

BTRPTDNAE B.taurus repeat region DNA.

LOCUS	BTRPTDNAE	482 bp	DNA	MAM	16-AUG-1993
DEFINITION	B.taurus repeat region DNA.				
ACCESSION	Z25529				
KEYWORDS	repeat region.				
SOURCE	cattle.				
ORGANISM	Bos taurus				
	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;				
	Theria; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae.				
REFERENCE	1 (bases 1 to 482)				
AUTHORS	Szemraj, J., Plucienniczak, G., Jaworski, J. and Plucienniczak, A.				
TITLE	Evidence for homological recombination with participation of the bovine alu-like sequences				
JOURNAL	Unpublished				
STANDARD	full automatic				
REFERENCE	2 (bases 1 to 482)				
AUTHORS	Plucienniczak, A.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-AUG-1993) PLUCIENNICZAK A., PP TERPOL, LABORATORY OF GENETIC ENGINEERING, P.O.W. 57, SIERADZ, POLAND, 98-200				
STANDARD	full automatic				
COMMENT	NCBI gi: 396758				
FEATURES	Location/Qualifiers				
source	1..482				
	/organism="Bos taurus"				
	/clone="pUJ3.24"				
	/dev_stage="calf"				
	/tissue_type="thymus"				
repeat_unit	133..482				
	/partial				
	/note="Truncated 5' part of BDDF."				
	/rpt_type=DISPERSED				
	/evidence=experimental				
	/rpt_family="Bovine Dimer Driven Family (BDDF)"				
	/label=BDDF				
	/citation=[1]				
repeat_unit	373..426				
	/partial				
	/note="5' part of bovine alu-like monomer."				

/rpt_type=FLANKING
/evidence=experimental
/rpt_family="bovine alu-like"
/citation=[1]

BASE COUNT 135 a 109 c 124 g 114 t
ORIGIN

Initial Score = 18 Optimized Score = 18 Significance = 5.15
Residue Identity = 85% Matches = 18 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

GGGTCGATGGTGGAGAGGTCGTGACGAGAATGTAGTCCACTGGAGAGGGGAATGGCAAACACTTTCAGTATT
 340 350 360 370 380 390 400

 X 10 20
 AAAAGTGCAAAAGCCTAGGAC
 ||||| ||||||| |||||

CTTGCCCTTGAGAACCCCATGAACGTATGAAAAGGGGCAAAAGCATAGGATAGCTGAAAGAGGGAACCCCCAGT
 410 420 430 X 440 450 X 460 470

CGATAGG
 480

> O (
OI IO IntelliGenetics
> O (<

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file jan11f.res made by on Wed 11 Jan 95 12:47:30-PST.

Query sequence being compared: CL26' (1-21)
Number of sequences searched: 302507
Number of scores above cutoff: 4881

Results of the initial comparison of CL26' (1-21) with:
Data bank : EMBL-NEW 10, all entries
Data bank : GenBank 85, all entries
Data bank : GenBank-NEW 10, all entries
Data bank : HIV-NA 7, all entries
Data bank : Issued_NA , all entries
Data bank : N-GeneSeq 16.3, all entries
Data bank : UEMBL 40_85, all entries
Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		
Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	6	7	3.01

Times:	CPU	Total Elapsed
	00:13:03.06	00:13:28.00

Number of residues: 276734581
Number of sequences searched: 302507
Number of scores above cutoff: 4881

Cut-off raised to 4.
Cut-off raised to 5.
Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.
Cut-off raised to 10.
Cut-off raised to 11.

Cut-off raised to 12.
 Cut-off raised to 13.
 Cut-off raised to 14.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 3 standard deviations above mean ****						
1. BTRPTDNAE	B.taurus repeat region DNA.	482	18	18	3.99	0
2. RABTCRGAM	Rabbit T-cell receptor gamma	147	17	17	3.65	0
3. Q77574	Human genome fragment. (Prefe	200	17	17	3.65	0
4. HSAACMHG	H. sapiens putatively transcr	200	17	17	3.65	0
5. ATTS1638	A. thaliana transcribed seque	274	17	17	3.65	0
6. TBILTA124	T.brucei mRNA for variant sur	1688	17	17	3.65	0
7. TBRVSG	T.brucei rhodensiense mRNA fo	1732	17	17	3.65	0
8. U01312	Streptococcus pyogenes JRS4 p	1823	17	17	3.65	0
9. S52562	LH-2=LIM/homeodomain protein	2072	17	17	3.65	0
10. HUMSWX167	Human chromosome X STS sWXD16	239	16	17	3.32	0
11. GCREG35	Galago Alu repeat type II, GA	245	16	16	3.32	0
12. HSA127WB5	H. sapiens (D1S505) DNA segme	319	16	16	3.32	0
13. NEUMTOLI2	N. crassa mitochondrial oli2	335	16	16	3.32	0
14. PLYORIA	Human polyomavirus BK (strain	375	16	16	3.32	0
15. M88810	CEL01E12 Caenorhabditis elega	394	16	16	3.32	0
16. S67523	early gene, late gene (contro	401	16	16	3.32	0
17. ATTS2283	A. thaliana transcribed seque	408	16	16	3.32	0
18. HS311VF9	H. sapiens (D5S662) DNA segme	414	16	16	3.32	0
19. PLYORIB	Human polyomavirus BK (strain	424	16	16	3.32	0
20. ATTS1882	A. thaliana transcribed seque	429	16	16	3.32	0
21. PVBECS522	Human papovavirus BK, Gardner	455	16	16	3.32	0
22. SYNECR530	BKV hybrid (tr-530) early tra	487	16	16	3.32	0
23. SYNECR532	BKV hybrid (tr-532) early tra	515	16	16	3.32	0
24. SYNECR531	BKV hybrid (tr-531) early tra	558	16	16	3.32	0
25. PVBECS501	Human papovavirus BK, Gardner	559	16	16	3.32	0
26. CEZMTTGP	Green turtle mitochondrion tr	620	16	16	3.32	0
27. Q58456	BK enhancer-adenovirus-2 late	642	16	16	3.32	0
28. Q54210	BK enhancer-adenovirus 2 late	642	16	16	3.32	0
29. HUMRPO	Human gene for ret proto-onco	678	16	16	3.32	0
30. ZEFTRANB	Danio rerio mRNA. Tc1-like tr	706	16	16	3.32	0

Query sequence being compared: CL26' (1-21)
 Number of sequences optimized: 4881

Results of the optimized comparison of CL26' (1-21) with:

Data bank : EMBL-NEW 10, all entries
 Data bank : GenBank 85, all entries
 Data bank : GenBank-NEW 10, all entries
 Data bank : HIV-NA 7, all entries
 Data bank : Issued_NA , all entries
 Data bank : N-GeneSeq 16.3, all entries

Data bank : UEMBL 40_85, all entries
 Data bank : VectorBank 9, all entries

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	14
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		
Initial scores to save	30	Alignments to save	30
Optimized scores to save	30	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	14	15	0.78
Times:	CPU	Total Elapsed	
	00:01:01.91	00:01:09.00	
Number of residues:	23291943		
Number of sequences optimized:	4881		

The scores below are sorted by optimized score.
 Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. BTRPTDNAE	B.taurus repeat region DNA.	482	18	18	5.15	0
2. MUSMA	Mouse mRNA for ORF.	7222	16	18	5.15	0
3. S92205	rna12+=pre-rRNA maturation [S	3587	15	18	5.15	0
4. ZEFTRAN	Danio rerio Tc1-like transpos	1205	16	18	5.15	0
**** 3 standard deviations above mean ****						
5. HSAAACMHG	H. sapiens putatively transcr	200	17	17	3.86	0
6. TBILTA124	T.brucei mRNA for variant sur	1688	17	17	3.86	0
7. ATTS1638	A. thaliana transcribed seque	274	17	17	3.86	0
8. TBRVSG	T.brucei rhodensiense mRNA fo	1732	17	17	3.86	0
9. S52562	LH-2=LIM/homeodomain protein	2072	17	17	3.86	0
10. HUMSWX167	Human chromosome X STS sWXd16	239	16	17	3.86	0
11. Q77574	Human genome fragment. (Prefe	200	17	17	3.86	0
12. U01312	Streptococcus pyogenes JRS4 p	1823	17	17	3.86	0
13. RABTCRGAM	Rabbit T-cell receptor gamma	147	17	17	3.86	0
14. T16193	IB3700 Homo sapiens cDNA 3'en	498	15	17	3.86	0
15. ZEFTRAN	Danio rerio Tc1-like transpos	1241	15	17	3.86	0
16. SSIS1139	S.salivarius insertion sequen	1717	15	17	3.86	0

17.	YSKSTE12X	Kluyveromyces lactis STE12 ge	2678	15	17	3.86	0
18.	CEZC84	Caenorhabditis elegans cosmid	38955	15	17	3.86	0
19.	CEZC84	Caenorhabditis elegans cosmid	38955	15	17	3.86	0
20.	CEZC84	Caenorhabditis elegans cosmid	38955	15	17	3.86	0
21.	M28728	Figure 1. (B) Sequences in wt	51	14	17	3.86	0
22.	Q38699	Oligonucleotide 7 to insert g	63	14	17	3.86	0
23.	SV4MNKR5	simian virus 40/african green	115	14	17	3.86	0
24.	HSBA7H052	H. sapiens partial cDNA seque	231	14	17	3.86	0
25.	SV4MNKR4	simian virus 40/african green	250	14	17	3.86	0
26.	SV4STA	Rhesus macaque polyoma virus	384	14	17	3.86	0
27.	SV4MNKR3	simian virus 40/african green	593	14	17	3.86	0
28.	SV4STA4	Rhesus macaque polyoma virus	694	14	17	3.86	0
29.	HUMRAB6A	Homo sapiens GTP-binding prot	740	14	17	3.86	0
30.	HSRAB6A	Homo sapiens GTP-binding prot	740	14	17	3.86	0

1. CL26' (1-21)

BTRPTDNAE B.taurus repeat region DNA.

LOCUS	BTRPTDNAE	482 bp	DNA	MAM	16-AUG-1993
DEFINITION	B.taurus repeat region DNA.				
ACCESSION	Z25529				
KEYWORDS	repeat region.				
SOURCE	cattle.				
ORGANISM	Bos taurus				
	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;				
	Theria; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae.				
REFERENCE	1 (bases 1 to 482)				
AUTHORS	Szemraj,J., Plucienniczak,G., Jaworski,J. and Plucienniczak,A.				
TITLE	Evidence for homological recombination with participation of the				
	bovine alu-like sequences				
JOURNAL	Unpublished				
STANDARD	full automatic				
REFERENCE	2 (bases 1 to 482)				
AUTHORS	Plucienniczak,A.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-AUG-1993) PLUCIENNICZAK A., PP TERPOL, LABORATORY OF				
	GENETIC ENGINEERING, P.O.W. 57, SIERADZ, POLAND, 98-200				
STANDARD	full automatic				
COMMENT	NCBI gi: 396758				
FEATURES	Location/Qualifiers				
source	1..482				
	/organism="Bos taurus"				
	/clone="pUJ3.24"				
	/dev_stage="calf"				
	/tissue_type="thymus"				
repeat_unit	133..482				
	/partial				
	/note="Truncated 5' part of BDDF."				
	/rpt_type=DISPERSED				
	/evidence=experimental				
	/rpt_family="Bovine Dimer Driven Family (BDDF)"				
	/label=BDDF				
	/citation=[1]				
repeat_unit	373..426				
	/partial				
	/note="5' part of bovine alu-like monomer."				

/rpt_type=FLANKING
/evidence=experimental
/rpt_family="bovine alu-like"
/citation=[1]

BASE COUNT 135 a 109 c 124 g 114 t
ORIGIN

Initial Score = 18 Optimized Score = 18 Significance = 5.15
Residue Identity = 85% Matches = 18 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

GGGTCGATGGTGGAGAGGTCGTGACGAGAATGTAGTCCACTGGAGAAGGGAATGGCAAACACTTTCAGTATT
 340 350 360 370 380 390 400

 X 10 20
 AAAAGTGCAAAAGCCTAGGAC
 ||||| ||||||| |||||

CTTGCCCTTGAGAACCCCATGAACGTATGAAAAGGGCAAAAGCATAGGATAGCTGAAAGAGGAACTCCCCAGT
 410 420 430 X 440 450 X 460 470

CGATAGG
 480

checked list
1/11/95
[signature]